

# The anomaly called psi: Recent research and criticism

K. Ramakrishna Rao and John Palmer

Institute for Parapsychology, Box 6847, College Station,  
 Durham, N. C. 27708

**Abstract:** Over the past hundred years, a number of scientific investigators claim to have adduced experimental evidence for "psi" phenomena – that is, the apparent ability to receive information shielded from the senses (ESP) and to influence systems outside the sphere of motor activity (PK). A report of one series of highly significant psi experiments and the objections of critics are discussed in some depth. It is concluded that the possibility of sensory cues, machine bias, cheating by subjects, and experimenter error or incompetence cannot reasonably account for the significant results. In addition, less detailed reviews of the experimental results in several broad areas of psi research indicate that psi results are statistically replicable and that significant patterns exist across a large body of experimental data. For example, a wide range of research seems to converge on the idea that, because ESP "information" seems to behave like a weak signal that has to compete for the information-processing resources of the organism, a reduction of ongoing sensorimotor activity may facilitate ESP detection. Such a meaningful convergence of results suggests that psi phenomena may represent a unitary, coherent process whose nature and compatibility with current physical theory have yet to be determined. The theoretical implications and potential practical applications of psi could be significant, irrespective of the small magnitude of psi effects in laboratory settings.

**Keywords:** clairvoyance; extrasensory perception (ESP); methodology; parapsychology; psi; psychokinesis (PK); replication; scientific method; telepathy

## 1. Introduction

There is a large and growing body of experimental literature devoted to the study of certain anomalous interactions that seem to involve psychologically meaningful exchanges of information between living organisms and their environment. We call these interactions *anomalous* because they appear to exceed somehow the capacities of the sensory and motor systems as these are presently understood. These interactions are collectively designated by the term *psi*. *Parapsychology* is that branch of science that makes a systematic study of psi anomalies. In other words, it is the business of parapsychology to find explanations of psi anomalies through scientific inquiry.

Psi is traditionally divided into various subcategories, each of which has been the subject of experimental research. For example, parapsychologists have been testing whether subjects can acquire information that is shielded from their senses (*extrasensory perception*, or ESP) and whether subjects can directly influence external systems that are outside the sphere of their motor activity (*psychokinesis*, or PK). Experimenters have also sought to differentiate forms of ESP, such as "telepathy" (ESP for another's thoughts) and "clairvoyance" (ESP for external objects and events). ESP is sometimes reported to be time-displaced, in that the information may relate to a past event ("retrocognition") or a future event ("precognition"). In practice, it has often proved difficult to isolate these forms of psi experimentally, and nowadays they tend to be defined operationally rather than theoretically (e.g., it is clairvoyance when you do not have someone sending the target).

Somewhat contrary to common usage, we are not using the term *psi* to imply that the anomalous interactions are necessarily "paranormal," but rather that no *adequate* conventional explanation of the interactions has yet been offered. Phrases stating or implying the "existence" of psi will be used somewhat informally to indicate that certain interactions have achieved this status.

The term *paranormal* has been a source of some confusion both within and outside parapsychology, and thus we feel that a few comments on the term are in order. Paranormal was first discussed in relation to psi by the philosopher C. D. Broad (1953; 1962; see also Braude 1979b), who defined psychical research (the earlier term for parapsychology) as "the scientific investigation of ostensibly paranormal phenomena" (Broad 1962, p. 3). Broad was careful to use the term "ostensibly paranormal," by which he meant phenomena that seem *prima facie* to conflict with one or more of what he referred to as the "basic limiting principles" of nature. These are not the same as the laws of nature, but rather a more fundamental set of assumptions that "we unhesitatingly take for granted as the framework within which all our practical activities and our scientific theories are confined" (Broad 1953, p. 7). For example, the assumption that "it is impossible for a person to perceive a physical event or a material thing except by means of sensations which that event or thing produces in his mind" (Broad 1953, p. 10) is a basic limiting principle that governs our way of acquiring knowledge. A case of ESP, therefore, would be *ostensibly* paranormal; it would be *genuinely* paranormal only when and if it could be shown to really conflict with one or more of the basic limiting principles. It is the task

of parapsychology, according to Broad, "to investigate ostensibly paranormal phenomena, with a view to discovering whether they are or are not genuinely paranormal phenomena" (Broad 1962, p. 5).

Although Broad's reasoning is sound, the term *paranormal* has led to some difficulties in practice. For example, as noted above, it is commonplace to find the terms *psi* and *paranormal phenomena* being used interchangeably, implying that parapsychology has no subject matter unless the paranormality of the phenomena is accepted in advance! A second and subtler difficulty is more directly related to the term itself. By stressing the conflict between potential "paranormal" explanations of *psi* and "normal" science, and at the same time failing to acknowledge that what constitutes normal science is historically relative (i.e., it can change from one historical period to the next), the term *paranormal* leaves the connotation that explanations that violate the basic limiting principles are unscientific in some fundamental sense. This, of course, is not true. If a "paranormal" theory of *psi* were someday to be confirmed, the practical consequence would be a redefinition of "normal" science to accommodate the new theory. In other words, the "paranormal" would become "normal," and the distinction would break down. A similar objection to the term has recently been raised by Paul Kurtz (1981), a well-known critic of parapsychology.

It is our view that potential explanations of *psi* that violate the basic limiting principles of nature are scientifically legitimate and, along with conventional explanations, should be entertained from the outset in our efforts to explain *psi* anomalies. Such explanations, unorthodox as they may be, are nonetheless worthy of consideration for the simple reason that *psi* anomalies seem to violate the basic limiting principles *prima facie*. Things are not always what they seem, but the possibility that they are should certainly be considered. Thus, the distinction to which *paranormal* refers is a valid one, even though the term itself is problematic. Recently, Palmer (1986b) has proposed a neutral term, *omega*, to identify potential explanations of *psi* that go beyond the basic limiting principles. Thus, "paranormal" explanations would be labeled "omegic." Despite our reluctance to introduce neologisms, we think in this case an exception may be justified.

## 2. Background

Like conventional psychology, experimental parapsychology grew out of a need to account for people's experiences in the "real world." The first major survey of such experiences was conducted under the auspices of the British Society for Psychical Research in the last century (Gurney et al. 1886/1970). More recently, a survey conducted by the National Opinion Research Center of the University of Chicago revealed that a majority of Americans thought they had experienced one or more psychic events in their lives (Greeley & McCready 1975). Similar results have been obtained in other surveys in the United States (e.g., Palmer 1979), Europe (e.g., Green 1960; Sannwald 1963; Haraldsson et al. 1977), and Asia (e.g., Prasad & Stevenson 1968). Palmer's survey further revealed that for many of those who

reported psychic experiences, these significantly influenced their feelings, attitudes, and decisions in other areas of their lives. Whatever the explanation of psychic experiences, they happen, they are common, and they are often important to people. For these reasons alone, they deserve serious attention from scientists involved in the study of human behavior and cognition.

Although some parapsychological research has directly examined the evidential value and characterization of these spontaneous psychic experiences (e.g., Hart 1954; Rhine, L. E. 1962; Schouten 1982), the bulk of the research has been experimental, and we will limit ourselves to the latter in this target article. The first major experimental investigation of *psi* was conducted at Stanford University by John Coover (1917). Sustained research, however, did not begin until 1927, when J. B. Rhine arrived at Duke University to work with William McDougall. With the publication of J. B. Rhine's (1934/1973) monograph *Extrasensory Perception*, a scientific claim for the existence of ESP was made. It gave the field "a shared language, methods, and problems" (McVaugh & Mauskopf 1976), and it provided "radical innovation and a high potential for elaboration" (Allison 1973, p. 39).

Rhine's procedure was to have subjects guess the randomized order of the cards in a deck containing five examples of each of five geometric symbols: a star, circle, cross, square, and wavy lines. By chance, the subject should get 5 correct out of the 25. Standard statistical techniques were used to determine the likelihood that any given number of hits was statistically significant. If the average number of correct guesses per run of 25 exceeded 5 to a significant degree, and acquisition of information by artifactual means such as sensory cueing and logical inference was ruled out, ESP was considered to have been demonstrated.

Using this methodology, Rhine (1934/1973) reported highly significant results, especially with five selected subjects who were tested repeatedly over a number of years. Prior to August 1, 1933, all subjects in the program had completed a grand total of 85,724 trials, with an average score of 7.1 hits per run.

The reaction of the scientific community to Rhine's claim was understandably cautious and critical. Subsequent to the publication of the monograph, there were 35 criticisms contained in 56 published reports. Some of these criticisms were specific and others were merely speculative. The specific criticisms had to do with Rhine's methods of data collection and statistical analysis. These criticisms and Rhine's responses are fully documented in the book *Extrasensory Perception After Sixty Years* (Rhine et al. 1930).

The first line of criticism dealt with the experimental conditions. One essential requirement for an acceptable ESP experiment was that data should be collected under conditions that provide no reasonable opportunity for sensory leakage of information or inferential knowledge of the targets. Skinner (1937), Woffle (1938), and J. L. Kennedy (1938), among others, pointed out that under certain lighting conditions the commercially produced ESP cards could be read through their reverse sides. Rhine responded that the original experiments were conducted with hand-printed ESP cards that were free from such defects and that in his more formal experiments

★

the use of screens and distance prevented the subjects from obtaining any visual cues from the cards. Kennedy (1938), Kellogg (1936), and Leuba (1938) argued that an increase in the experimental rigor of ESP research had resulted in a corresponding decline in ESP results, suggesting that extrachance ESP scores were due to loose experimental conditions. To this Rhine responded that his most rigorously controlled experiment, the Pearce-Pratt series, did give highly significant results (Rhine et al. 1940). Although this experiment was later challenged by critic C. E. M. Hansel (1966) – with questionable success (Hansel 1980; Rhine & Pratt 1961; Stevenson 1967) – as being susceptible to fraud on the part of the subject, it was still more rigorously controlled than the other experiments in the original data base and thus supported Rhine's point.

✓ The second line of criticism related to data analysis. Willoughby (1935), Kellogg (1936), Heinlein and Heinlein (1938), Herr (1938), and Lemmon (1939) criticized various features of the statistical analysis used by Rhine and his colleagues. In particular, the criticism focused on Rhine's assumption that the binomial theorem is applicable to "closed decks," decks in which the number of times each type of card appears is not free to vary. This aspect of the methodological debate essentially ceased in 1937, when Burton Camp, President of the Institute of Mathematical Statistics, stated that Rhine's "statistical analysis is essentially valid. If the Rhine investigation is to be fairly attacked it must be on other than mathematical grounds" (Camp 1937). For further details, see Burdick and Kelly (1977).

It would be wrong to conclude from this, however, that Rhine's experiments were perfect and that they had conclusively eliminated every alternative explanation. In retrospect, one could suggest improvements in the experimental conditions of his experiments. But for his time, Rhine's best experiments were ahead of others in the behavioral sciences. The experimental precautions he took, including two-experimenter controls and double-blind procedures, were rare in other disciplines at that time. Nonetheless, much of the early criticism of Rhine's experiments was helpful in progressively raising the standards of ESP research and reducing the possibility of experimental errors and artifacts.

Since the publication of Rhine's monograph over fifty years ago, there have been hundreds of experimental reports of evidence for psi. Yet skepticism has not decreased. Psi results are generally ignored in mainstream science, and when called to the attention of scientists they are apt to arouse suspicion. When specific criticisms are voiced, they generally include the following: (1) There is no "conclusive" experiment in parapsychology's long history; (2) there is no repeatable psi experiment; (3) the so-called significant psi results are disparate, incoherent, and isolated one-shot observations that do not merit scientific attention; (4) the results themselves are nonsensical in that they do not suggest any lawful relationships or progressive research programs; and (5) even if psi is real, it is too weak to be of any practical importance. If such perceptions were strongly supported by all the available data, it would be right to ignore parapsychology's claims. But the fact (as we hope to show in the following pages) is that (1) there are good experiments that seem to provide evidence for the existence of psi by reasonable standards

of judgment; (2) the results are repeatable at a respectable rate of replication; (3) the experimental observations in parapsychology are not unrelated, and significant patterns involving large bodies of experimental data are apparent; (4) a wide range of process-oriented research has focused on a single cognitive process that may be seen to give coherence and even a degree of consistency to a diverse array of experimental results; and (5) the small magnitude of most current psi effects is irrelevant to both their theoretical importance and their potential applicability.

### 3. The question of the "conclusive" experiment

Referring to parapsychology, Phillip H. Abelson (1978), Editor of *Science*, is quoted in *U.S. News and World Report* as saying that "extraordinary claims require extraordinary evidence." This statement implies that the strength of evidence required to establish a new phenomenon is directly proportional to how incongruent the phenomenon is with our prior notions. Our prior notions, however, are not always self-evident truisms. They are derived from, among other things, prevailing religious and cultural beliefs, personal experiences and observations, and our general world view. They are translated into subjective probability estimates and determine the evidential demands we make for a given claim. If the subjective probability of a disputed claim is zero, then no amount of empirical evidence will be sufficient to establish that claim. In serious scientific discourse, however, few would be expected to take a zero-probability stance because such a stance could be seen to be sheer dogmatism and the very antithesis of the basic assumption of science's open-endedness.

Nevertheless, the demand for extraordinary evidence of psi often seems to be derived from an implicit notion of its a priori impossibility. For example, some critics of psi research have demanded a "foolproof" experiment that would control for all conceivable kinds of error, including fraud by the experimenter(s). They have argued that if a claim is made for the existence of a phenomenon that conflicts with "established laws," it is much more parsimonious to assume error or even fraud on the part of the claimant than it is to assume the reality of that phenomenon (Price 1955; Hansel 1966). This argument is often identified with David Hume's (1825) maxim that "no testimony is sufficient to establish a miracle, unless the testimony be of such a kind, that its falsehood would be more miraculous than the fact which it endeavours to establish" (p. 115). Hume's maxim is a metaphysical statement, and it is inappropriate to use it when one speaks of empirical evidence. Moreover, his definition of a miracle as a universally nonexistent event is self-contradictory inasmuch as any claimed evidence in support of a miracle is also evidence against the universality of its nonexistence (Rao 1981a). As Saint Augustine remarked, "Miracles occur in contradiction not to nature, but to what is known to us of nature." It should also be kept in mind that Hume might not have regarded psi phenomena as miraculous or as anything more than extraordinary events.

The call for a totally "foolproof" study assumes that at a given time one can identify all possible sources of error

and how to control against them. Such a methodological stance is comparable to the epistemological position that one can determine for all time to come what is and is not possible. Again, the demand for experimental controls against experimenter fraud is unique to discussions of evidence for what are perceived to be extraordinary claims. Pushed to its extreme, the hypothesis of experimenter fraud becomes nonfalsifiable, in that it is impossible to be certain that fraud is completely eliminated in any given experiment.

The concept of a "conclusive" experiment, totally free of any possible error or fraud and immune to all skeptical doubt, is a practical impossibility for empirical phenomena. In reality, evidence in science is a matter of degree; the fact that one can concoct alternative explanations of a finding does not automatically render that finding evidentially worthless. Evidentiality must be assessed on a continuum and in relation to the plausibility of and the empirical support for the competing hypotheses. These considerations demand that a "conclusive" experiment be defined more modestly as one in which it is highly improbable that the result is artifactual. In this sense, we think a case can be made for "conclusive" experiments in parapsychology.

### 3.1. Schmidt's REG experiments

A defense of the existence of probabilistically conclusive parapsychological studies requires a detailed review and discussion of any experiments that might qualify. Because such a treatment must be rather lengthy, we will limit ourselves to a single group of experiments as an example. Although they are somewhat dated, we have chosen Helmut Schmidt's (1969a; 1969b) reports on random event generator (REG) experiments because (a) they represent one of the major experimental paradigms in contemporary parapsychology; (b) they are regarded by most parapsychologists as providing good evidence for psi; and (c) they have been subjected to detailed scrutiny by critics. In no sense do we imply that these are the only good experiments the field has to offer. Nor do we believe, for the reasons stated above, that there can be any crucial experiment or experimental program on which the case for psi does or could rest exclusively.

At the time of conducting these experiments, Helmut Schmidt was a physicist at Boeing Scientific Research Laboratories. The studies were designed to test the possibility of ESP and were carried out with the help of a specially built machine that seemed to rule out all artifacts arising from recording errors, sensory cues, and subject cheating. The safety features of the Schmidt machine are actually superior to those of the VERITAC machine used earlier by Smith and his colleagues to test for ESP (Smith et al. 1963). Hansel (1966) had praised VERITAC as "admirably designed" and had suggested that it could be "standardized for testing subjects for extrasensory perception" (p. 172).

The Schmidt machine randomly selected targets with equal probability and recorded both the target selections and the subject's responses. The subject's task was to guess which of four lamps would light and to press the corresponding button if he was aiming for high scores (or to avoid that button if aiming for low scores). As Schmidt (1969b) described it:

During a test, the subject sits in front of a small panel with four pushbuttons and four corresponding colored lamps. Each of the pushbuttons simultaneously activates a recorder switch and a trigger switch. The recorder switch serves to register which of the buttons has been pressed. The four trigger switches are connected in parallel such that pressing any one of the buttons closes a circuit, in turn triggering the random lighting of one of the four lamps. The system is designed so that on repeated pressing of the buttons the lamps light in random sequence, i.e., each lamp lights with the same average frequency, and there is no correlation between successively lit lamps or between the buttons pushed and the lamps lit. (p. 101)

Random lighting of the lamps was achieved, following the subject's response, by a sophisticated electronic random event generator that used a radioactive source, strontium 90. (See Schmidt [1970b] for a more complete account of the hardware design and methods of statistical evaluation.) The REG was extensively tested in control trials and found not to deviate significantly from chance.

The sequence of buttons pressed and lamps lit is recorded automatically on paper punch tape. In the research reported here, the two types of test (trying for a high or low number of hits) were recorded in different codes, such that the evaluating computer could distinguish between them. The number of trials made and hits obtained were displayed to the subject by electromechanical reset-counters. These numbers were also registered by nonreset counters, and the readings of all counters were regularly recorded by hand. This record agreed with the results obtained from the paper tape. The equipment was fraud proof, so that one could, in principle, let the subjects work alone. This was done, however, only in a small part of the tests with subject OC in the first experiment and did not increase the scores. In all other tests the writer was present in the same room with the subject. (Schmidt 1969b, p. 103)

Schmidt's first report was based on two experiments. The subjects in this study were preselected on the basis of their performance in the preliminary tests. In the first experiment there were three subjects. All of them attempted to obtain high scores. Together they did 63,066 trials and scored 16,458 hits, which was 691.5 more than mean chance expectation (MCE). The probability that such a result occurred by chance is smaller than  $2 \times 10^{-9}$ .

In the second experiment, two subjects from the first series and one new subject participated. One subject aimed for high scores and another for low scores. The third aimed high in some trials and low in others. The total number of trials was 20,000. Of these, 10,672 were high-aim trials and 9,328 were low-aim. The combined deviation of hits in the desired direction was 401 greater than MCE, which has an associated probability smaller than  $10^{-10}$ .

In the third experiment, Schmidt (1969a) tested six subjects, including two who had participated in the trials just described. The experiment was designed to test primarily for clairvoyance; the targets were digits from a random number table further shuffled by a congruential generator and recorded on paper punch tape. The subjects completed a total of 7,091 high-aim trials and 7,909

low-aim trials, for a grand total of 15,000. The combined deviation of hits in the desired direction was +260 ( $p = 3 \times 10^{-6}$ ).

## 2. Criticisms of Schmidt's REG experiments

Hansel (1980) discussed the "weaknesses" in Schmidt's experiments under three headings: (1) experimental design, (2) unsatisfactory features of the machine, and (3) inability to confirm the findings. He criticized the experimental design (a) for its failure to specify in advance the exact numbers and types of trials to be undertaken by each subject, (b) for its introduction of high-aim and low-aim conditions, and (c) for its lack of control of the experimenter.

Strictly speaking, criticism (a) is not relevant to the main purpose of the experiment, which was to determine not whether a given subject had ESP, but whether the experiment as a whole provided evidence for ESP. It is true, however, that in Schmidt's first experiment the number of *total* trials was also not specified precisely in advance. The high level of statistical significance obtained, however, renders the possibility that this factor could account for the results extremely unlikely. And, as Hansel acknowledges, this problem was corrected in the later experiments.

Criticism (b) is not substantiated. Noting that high-aim scores gave a positive deviation and low-aim scores a negative deviation, Hansel argued, "The fact that when positive and negative deviations are combined (maintaining their sign) they invariably give a purely chance score suggests that sampling from a common distribution may have taken place" (p. 230). In the first place, this argument fails to account for Experiment I, which involved only the high-aim condition and gave results that were just as significant as in the other experiments. Second, it is not clear how Hansel's criticism could apply to the other experiments, since the high and low conditions were assigned in advance and recorded automatically on paper punch tape in *different codes*. It would seem, in fact, that the introduction of high/low conditions has a certain additional merit in that one condition could be considered as a control for the other, as well as for machine bias. It is of interest that in discussing a different Schmidt experiment, Hansel (1981) himself criticized Schmidt for not having a control condition and recommended the introduction of a condition in which "the subject would not be 'willing' the light to move, or *he would aim at moving the light in the opposite direction*" (p. 32, our italics).

Hansel went on to contend that two different machines, one for high aim and the other for low aim, should have been used. But would not such a procedure have been criticized on the grounds that any obtained difference between the scores could have been due to the opposite bias of the two machines?

Criticism (c) is valid if by "control of the experimenter" Hansel meant control against experimenter fraud. It would have been entirely possible for Schmidt to fake the results if he had wished to. In the extreme case, for example, the whole experimental report could simply have been fabricated. We cannot conceive, however, how a nonintentional error on the part of the experimenter

could have artifactually produced the significant results.

Hansel's criticism (2) of the machine itself overlaps criticism (1-b) above and was discussed under that heading.

The final reason given by Hansel for his rejection of Schmidt's results was that they have not been confirmed. But this again seems erroneous, as will be shown in Section 4.1.1 below. Hansel made no mention of several experimental reports already in the literature that did in fact claim to confirm Schmidt's results; he instead referred only to the 1963 report of Smith et al., which gave null results when VERITAC was used to test for ESP. But even this comparison is problematic. First, the machines, experimental procedures, and manipulation of the psychological conditions differed markedly between the two studies. Second, Schmidt's subjects were carefully screened through pretesting procedures, whereas those who participated in the VERITAC experiment were not.

In a more recent publication, Hansel (1981) proposed a scenario that permits the *possibility* of trickery without providing any evidence that fraud had indeed occurred. Referring to one of Schmidt's experiments testing PK (Schmidt 1970a), he claimed that the subject could have shorted "either the +1 or the -1 input in the display panel to the earth line according to whether he wished to produce a high or a low score" (p. 30), which would account for the significant results. This argument seems fallacious. Because the REG and electronic counters were sealed in a metal box and the REG outputs were completely buffered, there was no way the subject could have tampered with the apparatus in the way Hansel suggests. Second, the data were independently recorded on punch tape. Had the subject shorted the tape machine, the total number of punches would have differed from the 128 specified for each run. Inspection of the tapes revealed no such discrepancies (Schmidt, personal communication).

Hansel went on to argue that the experimenter himself could have easily affected the punched record. This is debatable, but the possibility that Schmidt could have faked his data *somehow* has already been acknowledged. Recently, however, Schmidt has published a PK experiment designed to rule out the possibility of his (or his two co-experimenters) falsifying the data without collaboration from at least one of the others (Schmidt et al. 1986). Briefly, Schmidt, located at his lab in San Antonio, Texas, prepared lists of paired six-digit random numbers, called seed numbers, which were to be used to generate sequences of quasirandom binary digits by means of a complex mathematical algorithm known only to Schmidt. These seed numbers were mailed to the private address of Professor Luther Rudolph (L. R.) of Syracuse University. Robert Morris (R. M.) of the same university independently obtained a list of random target directions (high and low), one for each binary sequence, by using his laboratory's own REG. R. M. and L. R. exchanged their copies of the target-direction sequences and the seed numbers and then made the former available to Schmidt.

For the test proper, the subject in San Antonio entered the seed numbers into a computer. The computer then derived the binary sequences, which in turn governed the display on a computer screen of a pendulum swinging with random amplitude. The subject's task was to predict the



pendulum to swing with large amplitude on high-aim trials and with small amplitude on low-aim trials.<sup>1</sup> At the end of the run, which lasted for about a minute, the display showed the average swing over the run; thus the subject was given feedback about his rate of success.

Schmidt et al. reported significant results in support of their hypothesis. The combined Z for all the ten sessions was 2.71 ( $p < .005$ ). Because (a) the seed numbers for the binary sequences and (b) the target directions were independently derived by Schmidt and Morris, respectively, we know of no way Schmidt or Morris alone could have artifactually obtained the results. Such security procedures involving experimenters working independently in two different laboratories are seldom used in scientific research; but it is understandable that Schmidt felt that the validity of his results should not be based ultimately on his honesty alone.

Of course, the possibility of fraud is still not eliminated completely in this experiment. Even if we grant that Schmidt alone could not have faked the results, it remains possible, though less probable, that Schmidt and Morris, or Morris and Rudolph, could have conspired to produce them spuriously. Perhaps the logical next step is to have a critic participate as a co-experimenter, using the design of Schmidt et al. We would be curious to see how critics would react if such an experiment succeeded.

Hansel's criticisms of Schmidt's experiments are routinely taken as valid by most writers skeptical of psi (e.g., Alcock 1981). One of the few critics of psi who questions the basic premises of Hansel's reasoning on this point is Hyman (1981). "There is no such thing as an experiment immune from trickery," says Hyman. "Even if one assembles all the world's magicians and scientists and puts them to the task of designing a fraud-proof experiment, it cannot be done" (p. 39). Hyman, however, agrees with Hansel that Schmidt's PK experiments "do not provide an adequate case for the existence of psi" (p. 34). His principal reasons are twofold: (1) "Experience shows that the most promising research programs in parapsychology will most likely be passé within a generation or two" (p. 37); and (2) although Schmidt's randomization tests control against "long-term, or even temporary" machine bias, they do not "control against possible short-run biases in the generator output" (p. 38). He suggested, as did Hansel, that matched experimental and control sequences would have been a superior procedure.

The first point is not really a substantive criticism but merely counsels patience. The same thing can be said of research in some other areas of psychology. Moreover, "passé" does not necessarily mean "discredited," and much of the older research in parapsychology has withstood criticism rather well. The second point, as Hyman himself recognizes, "does not automatically provide an alternative explanation for how Schmidt obtained his results" (p. 38). Schmidt, who was aware of such a possibility, notes that "many more randomness tests were done than published to satisfy my own questions about the possibility of temporary random generator malfunctions" (Schmidt 1981, p. 41). Also, it is difficult to see how such malfunctions could account for subjects' ability to anticipate the timing and direction of the hypothesized short-run biases in Schmidt's early PK research, which used a high-aim, low-aim, protocol (Schmidt 1976). Finally, in some of Schmidt's more

recent work, direct comparisons *were* made between experimental and control sequences (e.g., Schmidt 1976).

#### 4. The question of replication

Even assuming that it was possible to determine conclusively the proper interpretation of a single experimental result, such an exercise would have little value in the context of doing science. The way the scientist functions is different from the way the historian does, for example. Unique events and isolated facts, unless they lead to, or are capable of leading to, some kind of general law, ordinarily hold little interest for science. Unlike historical facts, most phenomena of science are capable of being repeated. The Battle of Gettysburg will not be fought again. But psi as a laboratory effect must be reasonably capable of being observed repeatedly if one is to study it effectively and to understand it. Thus, as even Hansel (1980) concedes at one point, the importance of a fool-proof experiment recedes into the background as the phenomena become increasingly replicable.

Replicability does not necessarily mean that a finding must be reproducible on demand. It is not strictly an either-or situation, but a continuum (Rao 1981b). In this sense of statistical replication, an experiment or an effect may be considered replicated if a series of replication attempts provides statistically significant evidence for the original effect when analyzed as a series.

It may be argued that statistical replication is simply imperfect replication, and that a real phenomenon is something that is *in principle* repeatable. If a phenomenon has occurred once, it will occur again, provided the same set of circumstances arises. If one had perfect understanding of the critical variables, one could invariably predict its occurrence; if one had control over those variables one could produce the phenomenon on demand. The problem is that, in practice, perfect duplication of conditions is impossible to achieve. This is especially true in behavioral science experiments, where the causes of an effect are likely to be complex and difficult to pin down.

This does not mean that replicability cannot be improved substantially if some understanding of these crucial variables can be achieved. Indeed, such understanding is a major goal of scientific investigation. The other side of the coin, however, is that inquiry in such cases begins without this understanding. It is therefore inappropriate to demand absolute or even strong replicability of a phenomenon simply as a prerequisite for further research.

##### 4.1. Examples of replicability in parapsychology

Once we give up the notion of absolute replication, we can see that parapsychological phenomena are replicated in a significant statistical sense. For example, Palmer's (1971) review of so-called sheep-goat studies reveals that in 13 of the 17 experiments that used standard methods of analysis, the "sheep" (the subjects who believed in the possibility of ESP) obtained higher scores than did the "goat." The subjects who believed in the possibility of ESP, with 6 of the 13 achieving statistical significance. Carl Sargent's

non  
Seq

★

(1981) review of the reports published in English on the association between ESP and extraversion suggests that significant confirmations of a positive relationship occur at over six times the chance rate. However, the most extensive evidence for the statistical replicability of psi comes from the three data bases to be discussed in more detail below.

**4.1.1. REGs and psi.** Since the publication of the REG results discussed in Section 3.1 above, Schmidt has carried out several other successful REG experiments, mostly involving PK. More to the point, a number of other experimenters have successfully used the same devices or similar ones to test for psi.

The most prominent of these replications comes from the laboratory of Robert Jahn at Princeton University (Jahn 1982; Nelson et al. 1984). Jahn and colleagues use an REG based on a commercial electronic noise source. The hits are counted and displayed on the instrument panel and are permanently recorded on a strip printer as well as a computer. The subject's task is to influence the device mentally to produce an excess of hits on pre-designated PK+ trials and an excess of misses on PK- trials. In a total of 195,100 PK+ trials, 22 subjects obtained a mean score of 100.043 (MCE = 100). The mean for the same number of PK- trials was 99.965. Although small in magnitude, both these means are significantly different from mean chance expectation. The combined probability of the results is approximately  $3 \times 10^{-4}$ .

Each trial in Jahn's experiments incorporated alternate positive and negative counting on successive samples to provide an on-line internal control against any systematic bias in the noise source (i.e., positive and negative noise pulses alternated as hits). Also, baseline trials were recorded "under a variety of conditions before, during, and after the active PK trials" (Jahn 1982, p. 148) in a manner resembling that recommended by critics. The mean score for these 179,250 baseline trials was 100.005, which does not differ significantly from chance.

Radin et al. (1985) conducted a preliminary survey of all binary (two-choice) REG experiments published from 1969 (the year of Schmidt's first published REG experiment) to 1984. The sources sampled were the five major refereed parapsychological journals, the bound *Proceedings* of refereed papers presented at the annual Parapsychological Association Conventions, and a report of the Princeton data by Nelson et al. (1984), cited above. The reviewers defined an "experiment" as the "largest possible accumulation of data compatible with a single 'direction of effort' assigned to the subjects" (p. 205). In other words, data from all trials in which subjects aimed for the same binary outcome were pooled, ignoring other experimental conditions or classifications that may have pertained.

The reviewers uncovered 56 reports from approximately 30 principal investigators describing a total of 332 individual experiments. For 30 of the nonsignificant experiments, the authors of the reports provided insufficient data to allow the outcome (deviation of the hit total from chance) to be expressed quantitatively. In each of these cases, the reviewers randomly selected a Z-score from a normal (null) distribution of Z-scores to represent the outcome.

Seventy-one of the 332 experiments (21%) yielded

results significant at or beyond the 5% level (2-tailed), and the combined binomial probability for all the studies was  $5.4 \times 10^{-43}$ . The outcome was still significant, although more modestly so, when the data from Schmidt and the Princeton group were removed ( $p < 4.25 \times 10^{-7}$ ).

**4.1.2. Ganzfeld and ESP.** A second major research paradigm in which the replication rate over a relatively large number of studies has been systematically evaluated concerns ESP in the ganzfeld. The ganzfeld is a homogeneous visual field produced, for example, by placing a halved Ping-Pong ball over each eye with cotton filling around the edges. While the subject relaxes in a comfortable chair or bed, a uniform white or red light is focused on his face from about two feet. Sometimes the subject also listens to "pink" noise through attached earphones. Subjects typically report a pleasant sensation of being immersed in a "sea of light" (Honorton 1977, p. 459).

In a typical ganzfeld-ESP trial, the subject receives approximately 30 minutes of ganzfeld stimulation. After a period of adjustment and relaxation, the subject is asked to report all images, impressions, and so on, that occur at the time. From another room, an experimenter blind to the target monitors the subject's mentation via a microphone link and a one-way mirror. In a room located some distance from the subject, another experimenter acts as the agent. Some time after the subject has been in the ganzfeld, the agent-experimenter opens an envelope containing a target picture (randomly chosen from a pool of four), views it for about 15 minutes, and then stays in the room for an additional 10 minutes. After the completion of the ganzfeld period, the first experimenter gives the subject four pictures and asks him to assign them ranks of 1 through 4 for their correspondence to his mentation. At this time neither the subject nor the first experimenter knows which of the four pictures is the target. The agent-experimenter is then called in and reveals the target picture.

The first ganzfeld experiment in parapsychology was reported by Honorton and Harper (1974). The results of this experiment were subsequently replicated by Terry and Honorton (1976), Braud et al. (1975), and Sargent (1980), among others. According to a recent count adopted both by Honorton (1985) and critic Ray Hyman (1985b), there are 42 published ESP experiments that have used the ganzfeld procedure. After correcting for multiple analyses, if any, Honorton concluded that 19 of the experiments (45%) gave significant evidence for psi at or beyond the 5% level. Moreover, 26 of the 36 studies for which the direction of the effect could be clearly determined (72%) gave deviations in the positive direction, as compared to the 50% expected by chance. Hyman (1985b) dissented, concluding that the "rate of 'successful' replication is probably very close to what should be expected by chance given the various options for multiple testing exhibited in the data base" (p. 25). Later, however, he came to agree with Honorton that "there is an overall significant effect in this data base which cannot reasonably be explained by selective reporting or multiple analysis" (Hyman & Honorton 1986).

**4.1.3. The differential effect.** Another area of psi research with a large number of studies spanning a long period of time is the one dealing with the differential effect. This

the tendency of individual subjects to score differentially in successive ESP tests when these consist of two contrasting conditions, such as two different sets of targets or two different modes of response. In other words, subjects score above chance in one condition and below chance in the other. The first author's (K.R.R.'s) initial encounter with differential scoring occurred when he attempted to test subjects using both ESP cards and cards consisting of symbols to which the subjects were emotionally attached. In the first experiment, he found not only that the subjects obtained more hits than expected by chance with the cards of their chosen symbols, but also that their scores on cards with ESP symbols were lower than MCE. The scoring pattern with one set of cards was the mirror image of the pattern with the other (Rao 1962). Since then Rao has carried out a large number of tests under a variety of conditions and has found a rather consistent tendency on the part of subjects to show a bimodal response pattern when the ESP test consists of two contrasting conditions (Rao 1965).

It is interesting to note that evidence for the differential effect can be found in a number of studies carried out before and after Rao's studies, even when the experimenters themselves were not looking for it. For example, Rao and Krishna (in press) examined 72 independent comparisons between ESP scores obtained by the same subjects responding to two different classes of targets where interactions with other variables had not been predicted. Their sources were the five major refereed parapsychological journals and reports of refereed papers presented at Parapsychological Association conventions. They found that 45 of the 72 comparisons (63%) showed differential scoring, where we would expect 36 (50%) by chance ( $p < .05$ ). In 19 of the experiments (26%), the scoring rate between the two conditions was significantly different at or beyond the .05 level, though one would expect only 3.6 experiments (5%) to show significant differences by chance.

The meaning of the differential effect is not yet clear. It was not derived from a theory or model and provides no explanatory construct that might help us to understand psi. Rather, it reflects a characteristic of psi in a certain type of design, a characteristic that any adequate theory of psi must ultimately account for. One may call it a descriptive construct as distinct from an explanatory construct. Descriptive constructs are important in the early stages of scientific inquiry because, by defining what it is that a theory must explain, they serve to channel the process of theory development. Much of the research in modern parapsychology is directed toward identifying such descriptive constructs or "effects," with the objective of bringing closer to attainment the ultimate goal of a credible theory of psi.

**4.1.4. Overview.** The proportions of statistically significant studies in the three areas we have reviewed are as follows: REGs (21%); ganzfeld (45%); differential effect (26%). Given the expected success rate of 5%, these values are not trivial, and they compare favorably with comparable examples from psychology, such as the placebo effect (Moerman 1981) and the experimenter expectancy effect (Rosenthal & Rubin 1978). The latter authors, for example, reviewed evidence on the experimenter expectancy effect in eight types of experiments. The median replica-

tion rate was 39%. Except for one highly replicable topic (animal learning: 73%), the percentages ranged from 22% to 44%, which is very similar to what we find in parapsychology.

## 4.2. Some criticisms

A number of objections can be raised to the kind of procedure we have used in obtaining these replication rates, objections similar to those that have been raised in discussing experimenter expectancy effects (Barber 1969; 1973). Some of these objections will now be discussed in relation to the data under consideration.

**4.2.1. Comparability of studies.** One objection to such analyses is that the studies included are often not directly comparable. This objection has merit, but only to a point. We should not insist, for example, that all experiments be strict replications of one another. So long as they constitute conceptual replications, methodological differences can often be treated as random variables that actually serve to increase the generality of any conclusions that might be drawn from the analysis. On the other hand, it is usually desirable that the outcomes of the studies be represented by, or reduced to, some common metric. One of Hyman's (1985b) criticisms of the ganzfeld data base, for example, was that the studies used divergent and sometimes multiple measures of the dependent variable, and that the primary measure was sometimes not specified in advance. In response to this objection, Honorton (1985) computed a new analysis, using as a single, uniform measure Z-scores representing the proportion of trials in the experiment in which the subject correctly picked out the target during the judging (i.e., direct hits). This was the measure used in the original ganzfeld experiment by Honorton and Harper (1974), and it was the measure most frequently reported in the data base as a whole. Sufficient information for this analysis was provided for 28 of the 42 experiments in the data base. These experiments came from ten different laboratories. Twenty-three of the 28 experiments (82%) yielded positive Z-scores, 12 of which were individually significant at the .05 level on a one-tailed test. The cumulative Z-score for all 28 studies, computed by the Stouffer method (Rosenthal 1984), was 6.60 ( $p < 10^{-9}$ ).

Both Radin et al. (1985) and Rao and Krishna (in press) dealt with the uniformity issue in their analyses of the REG and differential effect experiments (discussed above) by using as a common metric Z-statistics. In the former case, these represented the proportion of trials that were hits; in the latter case, they represented the difference between the proportions of hits in the two conditions.

**4.2.2. Publication bias.** A second criticism concerns whether these analyses may suffer from biased selection and so-called publication artifact; that is, nonsignificant results may systematically go unreported, and therefore our sample of studies may not reflect the true state of affairs. Close scrutiny of the field suggests that publication bias cannot explain away the significant number of replications in parapsychology. Parapsychologists are sensitive to the possible impact of unreported negative results, more so than most other scientists. Our profes-

true  
population  
of samples



ional society, the Parapsychological Association (PA), has advocated a policy of publishing the results of all methodologically sound experiments, irrespective of outcome. Since 1976, this policy has been reflected in the publications of all the journals affiliated with the PA and in the papers accepted for presentation at the annual PA conventions.

This policy, however, cannot guarantee that researchers will submit negative findings for publication. Fortunately, thanks to a technique developed by Rosenthal (1979), we are able to estimate the number of unpublished and nonsignificant experiments that would be necessary to reduce an entire data base to nonsignificance. Honorton (1985), for example, used Rosenthal's technique to estimate that 423 nonsignificant ganzfeld studies would be needed to reduce the direct-hit studies in this data base to a nonsignificant level. Given the complex and time-consuming nature of the ganzfeld procedure, it is unreasonable to suppose that so many experiments exist in the "file drawer." As noted earlier, Hyman now agrees that selective reporting cannot account for the aggregate findings in the ganzfeld data base (Hyman & Honorton 1986).

A particularly ingenious way of estimating the extent of the file-drawer problem was implemented by Radin et al. (1985) in their analysis of the REG data base. By inspecting a graph of the distribution of outcomes, they noted a marked discontinuity at the Z-value associated with statistical significance: There were too many studies at the tail to make a smooth curve. They determined that the curve could be smoothed by adding 95 nonsignificant experiments to the data base. Doing this reduced the combined binomial probability of all the studies from  $5.4 \times 10^{-43}$  to  $3.9 \times 10^{-18}$ , still an impressive value. Using the Stouffer method, Radin et al. then estimated that ten parapsychology laboratories would each have needed to produce nonsignificant studies at the rate of 2.6 per month over the 15 years surveyed to cancel out the effect.

Finally, there are some areas in parapsychology where we can be reasonably certain we have access to all the experiments done. One such area concerns the relationship between ESP performance and the ratings obtained on the Defense Mechanism Test (DMT) developed in Sweden by Ulf Kragh and associates (Kragh & Smith 1970). Because the administration and scoring of this test requires specialized training available to only a few individuals, it has been possible for Dr. Martin Johnson of the University of Utrecht, the leading authority on the DMT and a man very sensitive to the issue of publication bias, to keep track of the number of relevant experiments conducted by qualified persons. In all ten of these studies the less defensive subjects scored higher on the ESP test. In seven of them, this effect was significant at the .05 level, one-tailed (Johnson & Haraldsson 1984).

**4.2.3. Controls and flaws.** A third line of criticism relates to experimental controls. It is argued, for example, that the replication of an experimental result by other experimenters "does not assure that experimental artifacts were not responsible for the results in the replication as well as in the original experiment" (Alcock 1981, p. 134).

It is true, of course, that the replication of an effect implies nothing directly about its cause. But it is also a basic premise of experimental science that replication

reduces the probability of *some* causal explanations, particularly those related to the honesty or competence of individual experimenters. As Alcock (1981) himself states in another context, "It is not enough for a researcher to report his observations with regard to a phenomenon; he could be mistaken, or even dishonest. But if other people, using his methodology, can independently produce the same results, it is much more likely that error and dishonesty are not responsible for them" (p. 133).

A more specific set of criticisms has been offered by Hyman (1985b) with reference to the ganzfeld-ESP data base. He concluded that the case for replication in this area is unconvincing because of the presence of methodological flaws such as potential sensory cues (e.g., including the target handled by the sender in the set given to the subject for judging), suboptimal randomization of targets (e.g., hand-shuffling), and multiple statistical analyses of the data. Honorton (1985) replied that Hyman made several unsupported assumptions in his analysis and interpretation of the ganzfeld-ESP data, and, in particular, that he often did not assign flaws properly with respect to his own criteria. Honorton presented his own analyses, arguing that the replication rate is not significantly influenced by the presence or absence of potential flaws in these studies. Although continuing to disagree on the seriousness of the "flaws," the reviewers have agreed that "the present data base does not support any firm conclusion about the relationship between 'flaws' and study outcome (Hyman & Honorton 1986). (Flaw analyses have yet to be reported on the REG and differential effect data bases.)"

The Hyman-Honorton ganzfeld debate is continuing in the *Journal of Parapsychology*. Whatever its final outcome, the discussion will lead to a more accurate interpretation of the data and better research in the future. In the final analysis, the case for psi cannot be won or lost by arguments over past experiments, but only by systematic and sustained new research that will survive the test of time. Honorton has recently reported continued success using an automated testing protocol that would appear to answer Hyman's methodological objections to the earlier ganzfeld research (Berger & Honorton 1985; Honorton & Schechter 1986).

**4.2.4. "Disbelievers" as replicators.** Several critics of psi research (Alcock 1981; Kurtz 1981; Moss & Butler 1978) have argued that the replication work must be done by investigators who are unsympathetic to psi, a category that would exclude most (but not all) parapsychologists. Moss and Butler, for example, argue that "replication by a qualified nonsympathetic observer is the only guard against results which may have been contaminated by a conscious or unconscious bias" (p. 1068).

We are now aware of its being common practice in other sciences to disqualify positive results from experiments conducted by researchers who are favorably disposed to the hypothesis they are testing. The personal beliefs of researchers are rarely reported and may often be difficult to determine reliably. We suspect, however, that if such a standard could be applied retrospectively to published research in psychology, for example, there would not be much left. The fact that parapsychologists are singled out for this treatment is symptomatic of the often ad hominem nature of the psi controversy. We have

yet to hear a critic suggest that negative results from "disbelievers" in psi be rejected on this basis.

Although it is reasonable to assume that experimenters who obtained strong positive results in the first few psi experiments they conducted were converted to a "belief" in psi by these results (if they were not "believers" already), we have far too few data to draw any conclusions about the distribution of attitudes of investigators at the time they undertook their first psi experiments. Thus we really do not know how many "disbelievers" have obtained positive psi results.

Finally, one cannot assume that confirmatory evidence, even from hardened "disbelievers," will necessarily be acknowledged as such. *BBS* readers might find it instructive in this connection to study what happened when certain members of the Committee for Scientific Investigation of Claims of the Paranormal quite unexpectedly confirmed Michel Gauquelin's astrological "Mars Effect." (See *Zetetic Scholar* 1982a; 1982b; 1983; and references contained therein.)

★ On the other hand, the fact that the outcomes of psi experiments seem to be sensitive, at least to a degree, to the identity of the experimenter or principal investigator is a legitimate cause for concern. This "experimenter effect" in parapsychology has long been recognized and extensively discussed within the field (e.g., Kennedy, J. E. & Taddonio 1976; White 1976a; 1976b); even some strong proponents of psi have had trouble obtaining positive results in their experiments. The jury is still out as to why this state of affairs exists. Until more is known, it is unwarranted and unfair to jump to the conclusion that the experimenter effect is due to fraud, negligence, or incompetence on the part of the successful experimenters, especially in the absence of supporting empirical evidence. The number of trained scientists who have obtained positive results in psi experiments is by no means inconsiderable, and many of these scientists have published in orthodox areas. More important, other plausible explanations of the experimenter effect can be proposed. For example, it is not implausible from a psychological point of view that an experimenter who does not expect positive results could convey this attitude to his subjects by nonverbal cues, thereby adversely affecting their confidence or motivation and thus their performance on the psi task. There is evidence from psychology for just such a process (Rosenthal & Rubin 1978). In addition, several studies within parapsychology that compared experimenters who had different attitudes or expectations about psi, or who behaved differently toward their subjects, have provided more direct support for this hypothesis (e.g., Honorton et al. 1975; Parker 1975; Taddonio 1976).

The correct explanation(s) of the experimenter effect can come only from more research. This will come sooner if more scientists outside the parapsychological community - "believers," "disbelievers," and neutrals - can be persuaded to undertake psi experiments of their own, and to publish their results irrespective of outcome. Despite our remarks earlier in this section, we think that the involvement of a wider range of investigators in psi research is important and we wish to encourage such involvement. Indeed, that was one of our objectives in writing this *BBS* target article. We and other parapsychologists would be pleased to consult with any quali-

fied scientist who would like to undertake such an experiment.

## 5. Patterns, order, and sense in parapsychology

Has parapsychology gone any further than merely suggesting that anomalies exist? We think it has. Although some work in the field is still concerned with demonstrating the integrity of the anomalies, emphasis in recent years has shifted strongly to so-called process-oriented research designed to uncover lawful regularities between psi and other psychological or physical variables. For example, there have been successful attempts to relate psi to subjects' beliefs and attitudes (Schmeidler & McConnell 1958), personality and motivation (Eysenck 1967; Honorton & Schechter 1986), and to cognitive variables such as memory (Rao et al. 1977), visual imagery (Kelly et al. 1975), and stereotypy of responses to ESP target sequences (Stanford 1975). We would like to focus here, however, on one hypothesis that appears to bring together a large and diverse body of experimental results: the idea that psi may be facilitated by procedures that result in the reduction of meaningful sensory and proprioceptive input to the organism, and the concomitant redirection of attention to internally generated imagery. This hypothesis is known in parapsychology as the noise reduction model.

Whatever its "real" mechanism, ESP may usefully be thought of as behaving like a weak signal that must compete for the information-processing resources of the organism. It follows that the reduction of ongoing sensorimotor activity may facilitate ESP detection by the organism. As illustrated in a book by the psychologist Harvey Irwin (1979), the noise reduction model fits in well with concepts that are widely accepted in cognitive psychology and information-processing theory. The model is particularly relevant to the notion of limits in the information-processing capacity of the organism (Kahneman 1973); namely, the more internal and external "noise" the system must process, the less is available to process possible psi information.

It is interesting that most of the traditional techniques of "psychic" development seem to involve some form of reduced vigilance or "noise reduction." For example, the practice of yoga, which is believed among other things to help develop ESP ability, appears to involve procedures that control habitual sensory, autonomic, and cognitive processes (Rao et al. 1978). The first five of the eight stages in Patanjali's yoga, for example, are preparatory and are aimed at achieving voluntary control of internal processes. The ability of yogins to exercise unusual control over heartbeat and EEG activity, to cause sweat on certain parts of the body, and become physiologically nonresponsive to external stimuli has been satisfactorily documented (Anand et al. 1961; Wallace 1970; Wallace et al. 1971). The final three stages of yoga are *dharana* (concentration), *dhyana* (meditation), and *samadhi* (a state of stillness of the mind). If the introspective accounts of the yogins are any guide, the *dharana* state seems to involve intense focusing of attention on a single object, whereas meditation (*dhyana*) enables the practitioner to hold that focus over an extended period of time, which is believed to result in a stand-still state of mind (*samadhi*).

his state is also described as an expansion of consciousness that goes beyond the object of perceptual attention (Dasgupta 1930). There is voluminous phenomenological information on this, along with a modicum of physiological data (see, e.g., Das & Gastaut 1955).

Historically, many of those who have claimed successful psi receptivity have also claimed that they did their best when they were physically relaxed and when the mind was in a "blank" state. Rhea White (1964), who reviewed the early literature on this topic, concluded that attempts "to still the body and mind" are common among the techniques used by successful psi subjects. Mary Sinclair, whom her husband, Upton Sinclair, found to be an excellent psi subject, recommended for a successful psi outcome that "you first give yourself a 'suggestion' to the effect that you will relax your mind and your body, making the body insensitive and the mind a blank" (Sinclair 1930, p. 180). White (1964) further elaborated this technique and classified it into four stages: (1) relaxation; (2) engaging the conscious mind by keeping it blank or focusing on a single mental image or feeling, perhaps following this by a "demand" that the psychic impression come; (3) waiting patiently for the impression to appear; and (4) assessing rationally if the impression is psychic.

There is also a large body of experimental evidence that procedures enabling a subject to limit extraneous sensory and proprioceptive input are conducive to the manifestation of psi. Much of this evidence has been comprehensively reviewed by Honorton (1977), so we will limit ourselves to a brief discussion of work in five areas — ganzfeld stimulation, hypnosis, relaxation, meditation, and dreams.

### 5.1. Ganzfeld and ESP

The research on ESP in the ganzfeld has already been discussed at some length. One additional point may be added that is particularly relevant to the present discussion: Those studies that assessed the self-reported effects of the ganzfeld on subjects' state of consciousness have generally found that the largest mean deviation scores from chance on the ESP test occurred among those subjects who claimed the greatest psychological effect from the manipulation (Palmer 1978; Sargent 1980).

### 5.2. Hypnosis and ESP

There is an extensive experimental literature on ESP and hypnosis. Fahler and Cadoret (1958), for example, tested college students in two formal experiments using a clairvoyance type of card-guessing task. In half of the trials the subjects were "under hypnosis" as they attempted to guess ESP cards screened from their view, and in the other half they guessed the targets while in a waking state. The order of testing was counterbalanced. In both experiments the subjects did significantly better in the hypnotic condition than in the waking condition.

In a careful review, Ephraim Schechter (1984) evaluated data from 25 experiments in which ESP performance was compared in hypnotic and control conditions. The results of 5 of these experiments are uninterpretable for a variety of reasons. Of the remaining 20 studies, 16 show higher scores for the hypnotic condition, with 7 of them

showing statistical significance. None of the four reversals are significant.

### 5.3. Relaxation and ESP

The most extensive work in this area has been carried out by William Braud. In one of the best designed of these studies (Braud & Braud 1974), 20 volunteer subjects were assigned randomly to "relaxation" or "tension" conditions. Those in the relaxation condition went through a taped, progressive-relaxation procedure (an adaptation of Jacobson's) before taking an ESP test, which was to guess the picture being "transmitted" by an agent in another room. The subjects in the other group were given taped, tension-inducing instructions before they did the same ESP test. Each subject's level of physical tension was assessed through electromyographic recordings and self-ratings. Both measures revealed a significant decrease in tension among the subjects in the relaxation group and a significant increase among those in the tension group; as predicted, the ESP scores of the subjects in the relaxation group were significantly above chance and significantly higher than those of the subjects in the tension group.

Although no formal meta-analyses have been conducted on this data base, our own informal survey uncovered 13 series from six researchers that have reported significant effects (two-tailed) favoring the facilitative effect of relaxation, and only one significant reversal using the same criteria.

### 5.4. Meditation and ESP

Studies investigating meditation and psi suggest a positive relationship between these two variables. Rao et al. (1978) reported three series of experiments with a total of 59 subjects who had various degrees of proficiency in yoga and meditation. The subjects were given two ESP tests both before and after they meditated for at least half an hour. In one test the subjects "blind matched" cards with ESP symbols against target cards concealed in opaque black envelopes, and in the other test they attempted to describe concealed pictures. The results of both tests yielded independently significant premeditation-to-postmeditation differences when the three series were pooled. The card-testing results were also significant for each of the three series separately.

Again, no formal meta-analyses have been conducted on this data base. However, our own informal survey uncovered 12 series from six researchers that have reported significant effects (two-tailed) favoring the facilitative effect of meditation, and only one significant reversal, using the same criteria.

### 5.5. ESP in dreams

Finally, mention should be made of a successful series of experiments on ESP in dreams conducted at Maimonides Medical Center (Ullman et al. 1973). In a typical experiment, a sender attempted to transmit the content of a randomly selected art print to a subject sleeping in an isolated room. When physiological monitoring indicated that the subject was dreaming, an experimenter blind to the target awakened the subject and elicited a dream report. The following morning, a tape of the dream

reports was played back to the subject, who added associational material and a "guess for the night." Subsequently, outside judges and/or the subject attempted to match the randomly ordered targets and dream transcripts from a series of sessions on a blind basis.

In an article that appeared recently in *American Psychologist*, Irvin Child (1985) reviewed 15 separate series from the Maimonides program. After eliminating data from analyses that may have been compromised by non-independence of the judgments, he concluded that the remaining data were collectively significant both for the independent judges and for the subjects as judges. Child's article also documents several instances of gross misrepresentation of the Maimonides experiments in commentaries by critics.

In contrast to the other research considered in this section, there have been no independent replications of the Maimonides research that have provided significant results. Two major failures to replicate have been reported (Belvedere & Foulkes 1971; Foulkes et al. 1972), and one other is equivocal (Globus et al. 1968).

### 5.6. Some criticisms

Considering the legendary elusiveness of psi, the rate of reported success in the psi studies involving sensory noise reduction, although far from perfect, is impressive, even more so because the results appear to make sense in the context of both traditional psychic training practices and theories from orthodox psychology. One could of course point out that studies such as the so-called remote-viewing experiments (Targ & Puthoff 1977), which do not involve any explicit procedures for reducing sensory noise, have also recorded success rates of about 50%, arguing that our rationale is unsupported by these studies. However, such an argument does not take into account the fact that most of the successful remote viewing experiments, unlike the experiments discussed above, used subjects that were preselected for psychic talent and thus less likely than ordinary volunteers to need a supportive cognitive state to perform successfully. Second, there is reason to believe that at least some of these subjects attempted to reduce noise on their own. Marilyn Schlitz, a highly successful remote viewing subject, put herself in a "calm state throughout," even though she used no formal relaxation procedure (Schlitz & Gruber 1980). Dunne and Bisaha (1978) asked their remote viewing subjects to "relax and clear their minds" prior to the remote viewing test.

Even if one were to concede that successful remote viewers are generally in an ordinary state of consciousness during the psi task, it does not follow that they might not have performed even better had they been in an altered state of the type we have been discussing. This observation, however, brings to light another criticism of the studies supporting the noise reduction model. Many of these studies, in particular most of the ganzfeld and relaxation experiments, failed to use control groups or other means of assessing whether the induction procedure was actually responsible for the positive scoring. Among those studies that did use such controls, the designs still did always preclude other interpretations of the results (see, e.g., Stanford 1987). Especially in the experiments using within-subjects designs, relative suc-

cess in the experimental condition might sometimes be attributable to expectancy effects or demand characteristics.

More research will be needed before the status of the noise reduction model can be conclusively determined. A large body of empirical data from diverse sources is nevertheless consistent with this hypothesis. This fact is sufficient to support the more modest point we are trying to make: Psi data fall into patterns that make psychological sense and encourage a systematic program of research.

## 6. Practical significance

The remaining criticism that needs to be addressed concerns practical significance. Even if one concedes that the preceding criticisms have been addressed satisfactorily, it can be argued that the results of psi experiments are trivial and of no practical or clinical importance. It is certainly true that the effect sizes in most psi experiments are small. For example, the effects reported by Schmidt in his REG experiments rarely exceed chance expectation by more than a few percent. Such outcomes hardly seem to be practically useful.

There are fallacies in this line of criticism, however. First, it fails to acknowledge the distinction between basic and applied research. Practical significance is indeed important if the objective is to determine whether a process can be applied to solve "real-world" problems. Parapsychology, however, is devoted almost exclusively to basic research, where the objective is to address theoretical issues. Psi results seem to violate expectations derived from generally accepted physical theory, and this makes them of theoretical interest irrespective of their magnitude. For example, many of the most important experiments in physics deal with effects of very small magnitude.

The above criticism is problematic even from the applied perspective, however, because techniques from information theory can be implemented to amplify a weak effect of the type commonly found in psi experiments. In one experiment, for example, Ryzl (1966) had the subject Stepanek guess whether the green or white sides of 30 cards placed inside opaque envelopes were uppermost. The cards were rerandomized and Stepanek guessed the order again. This process was repeated until Stepanek's distribution of guesses on each of 10 principal cards favored either green or white to a prespecified degree. Other criteria involving the other 20 cards also had to be met. The result was a single "majority vote" on each of the 10 principal cards. In each of five experiments, Stepanek's majority votes duplicated the target order of the 10 principal cards perfectly (100%), although his success rate on individual guesses was only 62%. Other examples of this approach have also been documented (e.g., Carpenter 1975; Puthoff 1985).

The reason that psi has not yet been applied on a broad scale has to do not with the size of the effects but with their unreliability, which (as discussed above) probably reflects our lack of understanding of the factors that affect performance on psi tasks. Uncovering these factors is a prime objective of modern parapsychological research.

If psi anomalies do in fact turn out to represent some

heretofore unrecognized and far-reaching ability to acquire information and manipulate the environment, and this ability could be brought under conscious control, the practical applications and potential benefits to mankind seem almost limitless. It is easy to put parapsychologists on the defensive by citing the slow progress that has been made to date in coming to grips with the anomalies. What such an approach overlooks is the importance of solving the admittedly unsolved puzzle that the anomalies represent. It seems to us that too many commentators on both sides of the psi controversy place excessive faith in what amounts to little more than speculations about the true nature of the anomalies. Only by continued research, preferably supported in a meaningful way by the scientific community at large, will the speculations turn into knowledge.

## 7. Conclusion

We find that the frequency of replications, especially with regard to the noise reduction hypothesis, indicates that we are indeed on the trail of something interesting. At the same time, we cannot totally rule out the possibility that we may yet discover a hidden artifact or set of artifacts that would provide a satisfactory conventional explanation of the results (and which might, in their own way, likewise prove interesting). Such an open approach, which is widely shared within the parapsychological community (Parapsychological Association 1986), is dictated by the anomalous nature of psi and the fact that there is still no verified theory of the mechanism(s) involved in psi interactions. Scientists working in this area must accordingly approach *all* hypotheses with an attitude of skepticism and must show a readiness to look at various alternatives (Palmer 1986a). Critics with a great deal of a priori skepticism about psi have reasonable grounds for not accepting omegic hypotheses – that is, that the anomalies represent a new principle of nature. At the same time, they have little justification for choosing to close their minds to the alternative possibility – namely, that the anomalies might reveal a currently unrecognized human capacity of great potential importance. If they do close their minds, they make the same mistake as those “believers in the paranormal” who refuse to study evidence and arguments contrary to their beliefs.

At the least, there is now an excellent *prima facie* case for the statistical repeatability of the anomalies under certain conditions. There appears to be a common thread running through these studies, diverse though they may be, in the techniques of eliciting and measuring psi. This commonality appears, at least in a crude and preliminary way, to make some theoretical sense and is leading to work now in progress at various laboratories to refine and consolidate the methods and concepts.

We have discussed here some experimental evidence

for the reality of psi, as well as the objections of critics to such evidence. We have also considered the idea that sensory noise reduction may be favorable to psi, sketching the experimental results that bear on this hypothesis. The following conclusions seem to emerge:

(1) Schmidt's results and many other parapsychological findings would be taken seriously if they related to a conventional area in science, for standard methodological and statistical criticisms have been answered.

(2) No single experiment, no matter how carefully designed and executed, can be expected to settle a controversial claim. The results of one good experiment do no more than make a claim. The significance of that claim is proportional to the degree that experiments supporting it are successfully replicated, and the degree of research and hypothesis-testing it generates. Also important is its potential for contributing to a theoretical understanding of the natural world and for practical application.

(3) The issue of replication and the meaning of experimental results in psi research have been a primary concern of parapsychologists. The discussion of the studies bearing on psi and sensory noise reduction and the rationale behind them show (a) a moderately significant rate of replication (in a statistical sense) and (b) the possibility of finding conditions that favor or inhibit psi. Together, these studies make a strong *prima facie* case for a genuine scientific anomaly and provide a viable research program.

(4) Further clarity and precision in the concepts and hypotheses are needed. Noise reduction, for example, needs to be defined more precisely. Some improvements in experimental design may have to be introduced to deal with the central issue of how psi operates. No mechanism or theory that would adequately explain psi has been validated. Those who accord an extremely low subjective probability to omegic hypotheses may therefore justifiably demand more and better evidence. But demanding such evidence is not the same as questioning the credibility of past research.

(5) The final settlement of the question of the status of psi will have to depend on further research. The scientific legitimacy of psi cannot be denied by personal innuendos and *ad hominem* arguments, just as it cannot be established by preaching. One can only hope that the climate of scientific opinion will be sufficiently tolerant to permit free and open inquiry by those who have the necessary skills and interest.

## NOTE

1. The theoretical rationale of the study was that the subject could psychokinetically influence the selection of the random seed numbers retroactively. We will not elaborate this hypothesis further, as it is not directly relevant to the control features of the experiment.